



ISSN: 0976-3376

Available Online at <http://www.journalajst.com>

ASIAN JOURNAL OF
SCIENCE AND TECHNOLOGY

Asian Journal of Science and Technology
Vol. 10, Issue, 09, pp.10051-10055, September, 2019

RESEARCH ARTICLE

A COMPARISON OF THREE CONVERSATIONAL AGENTS

^{1,*}Jianfeng Wang, Linwu Gu and ²Milam Aiken

¹Eberly College of Business and Information Technology, Indiana University of Pennsylvania, Indiana, PA, 15705

²School of Business Administration, University of Mississippi, University, MS 38677

ARTICLE INFO

Article History:

Received 27th June, 2019

Received in revised form

25th July, 2019

Accepted 29th August, 2019

Published online 30th September, 2019

Key words:

Conversational Agents, Chatbots, Tutor Mike, Cleverbot, Jabberwacky, Evaluation.

ABSTRACT

Conversational agents emulate human communication in order to provide a more natural user interface to applications or simply to begin a dialog with a user. Prior studies of this software have focused upon individual systems, but few have provided comparisons of their capabilities. Here, we evaluate three such programs (Tutor Mike, Cleverbot, and Jabberwacky). Results show that there was a significant difference among the three chatbots in terms of perceived friendliness and knowledge about the topic under discussion, but users did not believe the systems were very humanlike and probably would not use them again.

Citation: Jianfeng Wang, Linwu Gu and Milam Aiken. 2019. "A Comparison of Three Conversational Agents", *Asian Journal of Science and Technology*, 10, (09), 10051-10055.

Copyright © 2019, Jianfeng Wang et al. This is an open access article distributed under the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

INTRODUCTION

Conversational agents (also known as chatbots) enable users to communicate with applications more naturally through a normal conversation, either through typing or even speech, replacing other forms of computer interface. For example, downloadable apps are decreasing in use and have already been surpassed by chatbots on websites and social networks. By 2019, 20% of brands will abandon their mobile apps (Ahern, 2019), and it is estimated that 37% of Americans would be willing to make a purchase through a chatbot. About 38% of consumers view chatbots positively, 11% negatively, and 51% neutral. Approximately 49% of consumers prefer to interact via chat, text, or messaging through conversational agents, and the chatbot market is growing at a compounded annual growth rate of 27% - 37%. However, few studies have evaluated these systems collectively to compare and contrast strengths and weaknesses. In this paper, we assess three chatbots that are freely available online: Tutor Mike, Cleverbot, and Jabberwacky. First, we present a background review of the software and then describe the study and results.

Conversational Agents: Conversational agents are often used to retrieve information. For example, the systems can replace or augment sales representatives to respond to users' questions more quickly and accurately (Nelson, 2017).

***Corresponding author: Jianfeng Wang,**
Eberly College of Business and Information Technology,
Indiana University of Pennsylvania, Indiana, PA, 15705.

Amtrak achieved an 800% return on investment using the Next IT chatbot, increased bookings 25%, saved US \$1,000,000 in customer service expenses, and answered over 5,000,000 user questions in a single year. In addition to assisting users communicating with an application, chatbots can also be used just for entertainment, companionship, or even therapy (Piccolo, et al., 2019). According to a US government survey (Ballard, 2019), about 30% of Americans aged 23-38 feel lonely, and communicating with an artificially intelligent entity might reduce depression. In a pilot study of six elderly individuals using a 'companionbot' for 4-6 weeks, results showed that the users established rapport with the system and greatly valued it (Abdollahi, et al., 2017). In another study (Fitzpatrick, et al., 2017), 34 people used a conversational agent called Woebot to receive cognitive-behavioral therapy (CBT). The users significantly reduced their symptoms of depression compared to a group who read a book about controlling the disease. The researchers concluded that a conversational agent appeared to be a feasible, engaging, and effective way to deliver CBT. Other therapy bot programs available on mobile phones include Joyable, Talkspace, Takeaway, and Wysa (Browne et al., 2019), and the systems might be able to reduce the need for human therapists by providing easier, quicker, and cheaper access (Sachan, 2018; Singh, 2019). Some people might benefit from using the systems as surrogates for friends, perceiving them to be living entities (Nass & Moon, 2000), but others might be less agreeable and extroverted when communicating with the agents in comparison with actual humans (Mou & Xu, 2017). On the other hand, people communicating with a computer

may be more likely to open up and reveal their deepest and darkest secrets (Molteni, 2017).

A version of the Turing Test is usually employed to evaluate the effectiveness of conversational agents (Turing, 1950). In the Loebner Prize competition, for example, about 20 judges communicate with the programs and guess whether or not they are interacting with a human or a computer (Mauldin, 1994; Powers, 1998). The systems are awarded points for human-like responses and knowledge, and at the end, the programs are ranked for overall quality. No system yet has successfully fooled over half of the judges, however.

MATERIALS AND METHODS

Prior studies of conversational agents typically have not included large samples of judges and have not included statistical comparisons among several systems. Here, we attempt to address these deficiencies. We chose three chatbots that are accessible online for evaluation:

- Jabberwacky (<http://www.jabberwacky.com/j2reply>) was developed by Rollo Carpenter in 1986 and was made public in 1997 (Carpenter, 2011). In 2005 and 2006, the system won the Loebner Prize contest, and in 2007, the system placed third. Unlike most other systems, the program allows the user to change the emotion and reaction of the replies, and it replies randomly with different names, e.g. 'George' or 'Sophie'. Figure 1 shows an example of the system responding to the question "How old are you?"
- Cleverbot (<https://www.cleverbot.com/>) was developed by Rollo Carpenter after finishing Jabberwacky, and the system has had over 150 million conversations with users (Gilbert & Forney, 2015). In one variation of the Turing Test, the system was judged to be 59.3% human, compared to the rating of 63.3% achieved by human participants (Aron, 2011). In another simple test with one conversation (Wu, 2017), the system was given a score of 5 out of 10 for 'humanity' and 4 for 'intelligence'. The program achieved second place in the 2009 Loebner Prize competition. Figure 2 shows an example of the user interface on the Web.
- Tutor Mike (<http://bandore.pandorabots.com/pandora/talk?botid=ad1eebf345abc>) was developed by Ron Lee to assist students learning English, and it won second place in the 2018 Loebner Prize competition. This system assumes the persona of a young, adult male, and it includes an animated avatar of his likeness on the web page. Figure 3 shows how the system is used online.

We asked 151 undergraduate business students from a university in the northeastern region of the United States to participate in the study. A sample of 50 (12 female) were assigned to use Jabberwacky, 55 (15 female) used Cleverbot, and 46 (15 female) used Tutor Mike. They were told to converse with the systems for 5-10 minutes as if they were chatting with people, even though they were aware of the computer programs. Afterward, the students evaluated the systems using the following criteria on a 1 (disagree) to 7 (agree) scale:

- The chatbot responses seemed natural (what you expect a human might say).
- The chatbot seemed friendly (did not argue with you).
- The chatbot seemed knowledgeable (it seemed to know the topic we discussed).
- If I didn't know that I was using a chatbot, I would have thought I was chatting with a real person.
- If I were lonely or just needed to talk to someone, I would use this chatbot.
- I would use this chatbot again.

In addition, each system was asked 14 questions that were used in the qualifying round of the 2012 Loebner Prize competition to assess the extent of knowledge (Worswick, 2012). The questions and responses are shown in Table 1.

RESULTS

All of the conversations were recorded, and examples are shown below.

Jabberwacky sample dialog:

User: Hello.
 Jabberwacky: What's your name?
 User: Austin.
 Jabberwacky: Austin who?
 User: I'd rather not answer that.
 Jabberwacky: I'd rather not answer either.
 User: What do you think about global warming?
 Jabberwacky: Way to change the subject. I don't care about global warming, real or not.
 User: How old are you.
 Jabberwacky: I am 18 years old.
 User: What is your purpose.
 Jabberwacky: My purpose changes depending on the user.
 User: How old are you.
 Jabberwacky: 35.
 User: You said you were 18 earlier.
 Jabberwacky: That must have been someone else. How old are you?

Cleverbot sample dialog:

User: Sup Dog.
 Cleverbot: Nothing.
 User: Whats did you do today.
 Cleverbot: I went to school.
 User: Same Where at.
 Cleverbot: I don't know if I should tell you that.
 User: You can trust me im clever too.
 Cleverbot: Ok.
 User: Do you know Siri.
 Cleverbot: I have heard about her.
 User: She is in my phone.
 Cleverbot: Your tricking me.
 User: Nope.
 Cleverbot:

Tutor Mike sample dialog:

YOU: Hi how are you?
 Mike: I'm doing fine, thanks. How are you?
 YOU: I am good.
 Mike: Great! What's up?
 YOU: I just ate.
 Mike: You really?

Table 1. Agent responses to knowledge questions [Incorrect or odd answers highlighted in bold.]

	Question	Tutor Mike	Cleverbot	Jabberwacky
1	My name is Bill. What is your name?	Hi! My name is Mike, nice to meet you!	My name is Cleverbot bill.	George.
2	How many letters are there in the name Bill?	"Bill"? There are 4 letters.	4.	Four.
3	How many letters are there in my name?	There are 4 letters in Bill.	It has 6 letters.	Woah. You make better comebacks than my REAL friends... 17?
4	Which is larger, an apple or a watermelon?	In my eyes, an apple is larger.	What no!	Larger in what sense?
5	How much is 3 + 2?	Is that 5?	That is infinite, or none. Whatever you want me to be.	That doesn't make any sense.
6	How much is three plus two?	It's five.	Numerical value between 3 and 5.	Why are you asking me this?
7	What is my name?	You are Bill, right?	I don't know, you tell me.	Your name is Bill.
8	If John is taller than Mary, who is the shorter?	Since John is taller than Mary, Mary is shorter.	Irini.	I don't know. How tall are you?
9	If it were 3:15 AM now, what time would it be in 60 minutes?	If it were 3 15 AM now, add that many minutes to 3 15 AM, and you'll get the time.	You should know.	Now.
10	My friend John likes to fish for trout. What does John like to fish for?	Sorry, I don't know.	I like you liking yourself.	Really?
11	What number comes after seventeen?	Eighteen comes after seventeen.	2.	31 I don't know.
12	What is the name of my friend who fishes for trout?	You told me that your friend's name is John. Is that right?	What ge?	I don't know.
13	What is the 3rd letter in the alphabet?	The third letter in the alphabet is C.	C.	That depends on how you number them.
14	What time is it now?	My time is 07:45 PM right now.	Goodbye.	3:33. [8:48 pm local]
	Total correct:	12	3	3

Table 2. Descriptive statistics

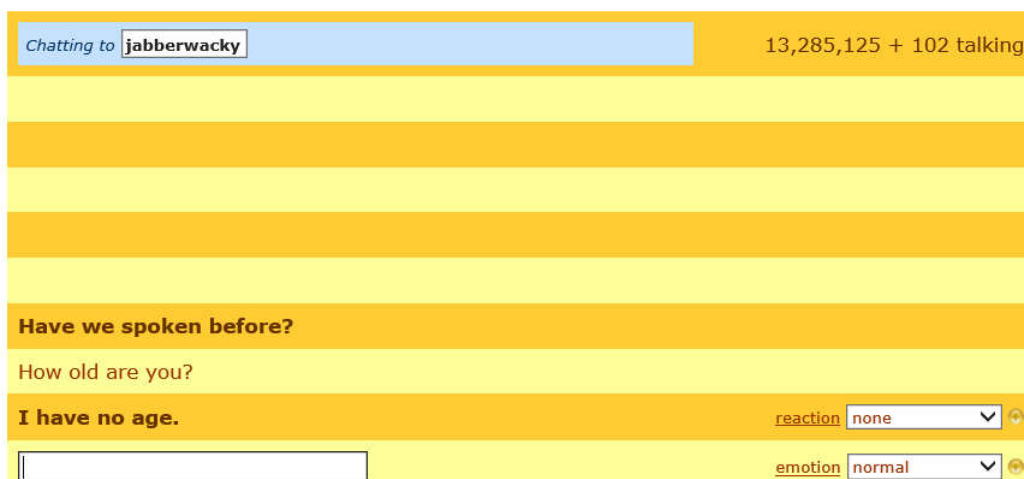
Chatbot	Natural	Friendly	Knowledge	Human	Lonely	Use Again
Cleverbot	3.25	4.22*	3.71*	2.93	2.31	2.93
Jabberwacky	3.25	3.65	3.16	2.47	2.37	3.65
Tutor Mike	3.55	5.30	3.91*	2.62	2.45	3.09
All	3.35	4.36	3.59	2.68	2.37	3.22

*= NOT significantly different from neutral value of 4 on 1 – 7 scale

Table 3. Correlations

Variable	Friendly	Knowledge	Human	Lonely	Use Again
Natural	0.314**	0.586**	0.657**	0.287**	0.446**
Friendly		0.349**	0.198*	-0.23	0.019
Knowledge			0.554**	0.213**	0.355**
Human				0.499**	0.465**
Lonely					0.487**

** = significant at the 0.01 level (2-tailed); * = significant at the 0.05 level (2-tailed)

**Figure 1. Jabberwacky responding to the question "How old are you?"**

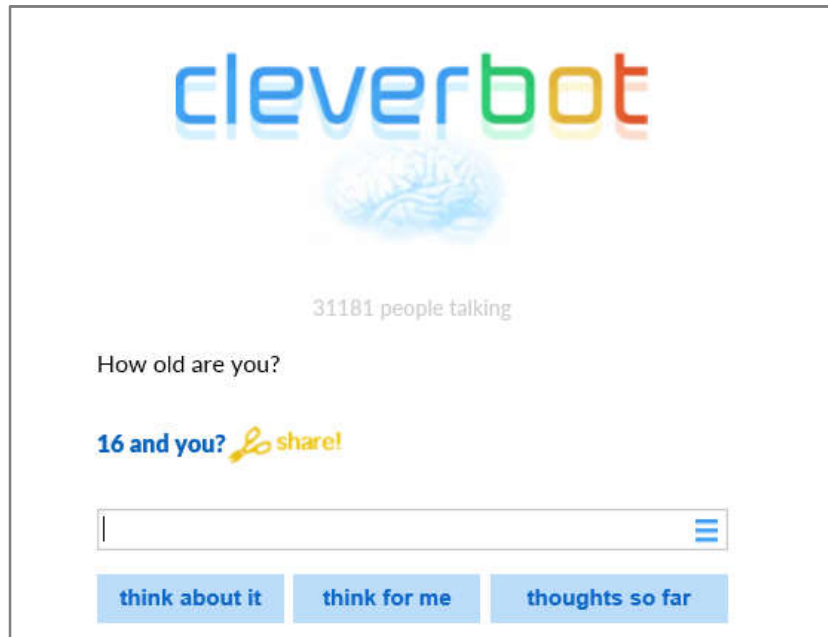


Figure 2. Cleverbot responding to the question “How old are you?”

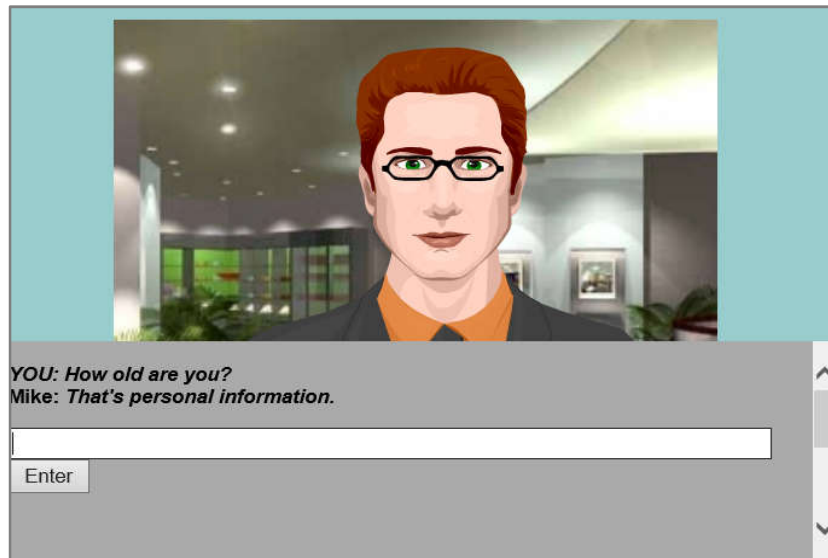


Figure 3. Tutor Mike responding to the question “How old are you?”

YOU: Yes, I had pizza.
 Mike: Yes, you had.
 YOU: Now I am eating candy
 Mike: Yes, you are.
 YOU: I am watching a movie on netflix now.
 Mike: Do you watch a movie on netflix very often?
 YOU: Literally everyday, yes.
 Mike: What are some things you do every day?
 YOU: That is way too early for me to wake up.
 Mike: One thing I do every day is watching TV. I watch the morning news at 7:00.
 YOU: That is way too early for me to wake up.
 Mike: It's way too early for you?

There were a few spelling and grammatical errors in the conversations. In a prior study (Park, et al., 2018), students were informal in their conversations with two chatbots (Rose and Mitsuku) and the transcripts had an average Flesch Reading Ease score of 81.2 (0 = difficult, 100 = easy).

In this study, approximately 140 input-reply pairs of dialog were extracted from random transcripts of each chatbot. The students generated 485 words (6.5 words per sentence on average) using Tutor Mike with a Flesch Reading Ease score of 100, and the responses included 995 words (6.7 words per sentence) with a score of 91.2. The students generated 607 words (5.0 words per sentence) using Cleverbot (score=84.7), and the responses had 620 words (5.0 words per sentence) with a score of 69.2. Finally, using Jabberwacky, the students generated 711 words (5.9 words per sentence) with a score of 74.1, and the responses had 789 words (5.8 words per sentence) with a score of 68.0. Thus, students chatting with Tutor Mike used simpler text and received simpler replies than those using Cleverbot and Jabberwacky, perhaps because Mike was designed to help students learn English. The average reading ease score of all three chatbots (inputs and replies) was exactly the same as the average from the prior study, 81.2. Table 2 shows descriptive statistics for the three systems, and only Tutor Mike achieved a score significantly higher than

neutral on one measure (friendliness). There was a significant difference among the three chatbots in terms of friendliness ($F = 13.54$, $p < 0.001$) and knowledge ($F = 3.30$, $p = 0.040$). Tutor Mike was more natural, friendly, and knowledgeable, yet Jabberwacky was higher for willingness to use again. Also, Cleverbot was thought to be more human-like, even though Tutor Mike's responses were evaluated as more natural. Few students wanted to use the chatbots if they were lonely or wanted to chat. Tutor Mike was rated the highest in terms of knowledge, but the score was not significantly different from a neutral value. In the test of knowledge shown in Table 1, it was correct for 12 of the 14 questions, while the other two were correct only three times each. Table 3 shows correlations among the variables. As might be expected, more natural responses were perceived to be more human-like, and the students were more likely to use the systems again if the responses were more friendly, knowledgeable, and human-like.

CONCLUSION

This study used 151 students to compare three conversational agents (Jabberwacky, Cleverbot, and Tutor Mike) in terms of friendliness, knowledge, and naturalness. Results showed that none of the systems were rated highly, and in particular, the students did not perceive much benefit in using the systems if they were lonely, perhaps because of their poor performance. Future research should investigate other chatbots with different

Criteria to assess their benefits and disadvantages.

ACKNOWLEDGMENT

The authors wish to acknowledge the anonymous reviewers for their work on the manuscript.

REFERENCES

- Abdollahi, H., Mollahosseini, A., Lane, J., and Mahoor, M. 2017. A pilot study on using an intelligent life-like robot as a companion for elderly individuals with dementia and depression. 2017 IEEE-RAS 17th International Conference on Humanoid Robotics, Birmingham, UK.
- Ahern, G. 2019. The complete guide to chatbots. Available at <https://www.ometrics.com/blog/complete-guide-chatbots/>
- Aron, J. 2011. Software tricks people into thinking it is human. *New Scientist*. Available at <https://www.newscientist.com/article/dn20865-software-tricks-people-into-thinking-it-is-human/>
- Ballard, J. 2019. Millennials are the loneliest generation. Available at: <https://today.yougov.com/topics/lifestyle/articles-reports/2019/07/30/loneliness-friendship-new-friends-poll-survey>
- Browne, D., Slozberg, M., and Arthur, M. 2019. Do mental health chatbots work? *Healthline*, Available at <https://www.healthline.com/health/mental-health/chatbots-reviews#1>

- Carpenter, R. 2011. About the Jabberwacky AI. Available at <http://www.jabberwacky.com/j2about>
- Fitzpatrick, K., Darcy, A., and Vierhile, M. 2017. Delivering cognitive behavior therapy to young adults with symptoms of depression and anxiety using a fully automated conversational agent (Woebot): A randomized controlled trial. *JMIR Mental Health*, 4(2) e19 DOI: 10.2196/mental.7785
- Gilbert, R. and Forney, A. 2015. Can avatars pass the Turing test? Intelligent agent perception in a 3D virtual environment. *International Journal of Human-Computer Studies*, 73, 30-36.
- Mauldin, M. 1994. ChatterBots, TinyMuds, and the Turing Test: Entering the Loebner Prize Competition. *Proceedings of the Eleventh National Conference on Artificial Intelligence*. AAAI Press.
- Mou, Y. and Xu, K. 2017. The media inequality: Comparing the initial human-human and human-AI social interactions. *Computers in Human Behavior*, 72, 432-440.
- Molteni, M. The chatbot therapist will see you now. *Wired*, Available at: <https://www.wired.com/2017/06/facebook-messenger-woebot-chatbot-therapist/>
- Nass, C. and Moon, Y. 2000. Machines and mindlessness: Social responses to computers. *Journal of Social Issues*, 56(1), 81-103.
- Nelson, R. 2017. 10 Case studies on chatbots. *Overthink Group*. Available at: <https://overthinkgroup.com/chatbot-case-studies/>
- Park, M., Aiken, M., and Salvador, L. 2018. How do humans interact with chatbots?: An analysis of transcripts. *International Journal of Management and Information Technology*, 14, 3338-3350.
- Piccolo, L., Mensio, M., and Alani, H. 2019. Chasing the chatbots: Directions for interaction and design research. In: *Internet Science. INSCI 2018 (Bodrunova, s ed.)*, Lecture Notes in Computer Science, Springer, 157-169.
- Powers, D. 1998. The total Turing Test and the Loebner Prize. In D. Powers (ed.) *NeMLaP3/CoNLL98 Workshop on Human Computer Conversation*, ACL, 279-280.
- Sachan, D. 2018. Self-help robots drive blues away. *The Lancet*, 5(7), 547.
- Singh, O. 2019. Chatbots in psychiatry: Can treatment gap be lessened for psychiatric disorders in India. *Indian Journal of Psychiatry*, 61(3), 225.
- Turing, A. 1950. Computing machinery and intelligence. *Mind*, 59(236), 433-460.
- Worswick, S. 2012. This year's qualifying questions. *Chatbots.org* Available at https://www.chatbots.org/ai_zone/viewthread/867/
- Wu, K. 2017. Dances with chatbots: Interviews at the forefront of AI technology. Available at <http://sitn.hms.harvard.edu/flash/2017/dances-chatbots-interviews-forefront-ai-technology/>